The Plant Cell, Vol. 23: 4266–4279, December 2011, www.plantcell.org

Identification of Plant *RAD52* Homologs and Characterization of the *Arabidopsis thaliana RAD52*-Like Genes [W]

Aviva Samach,[a] Cathy Melamed-Bessudo,[a] Naomi Avivi-Ragolski,[a] Shmuel Pietrokovski,[b,1] and Avraham A. Levy[a,1,2]

[a] Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[b] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

RADiation sensitive52 (RAD52) mediates RAD51 loading onto single-stranded DNA ends, thereby initiating homologous recombination and catalyzing DNA annealing. RAD52 is highly conserved among eukaryotes, including animals and fungi. This article reports that RAD52 homologs are present in all plants whose genomes have undergone extensive sequencing. Computational analyses suggest a very early *RAD52* gene duplication, followed by later lineage-specific duplications, during the evolution of higher plants. Plant RAD52 proteins have high sequence similarity to the oligomerization and DNA binding N-terminal domain of RAD52 proteins. Remarkably, the two identified *Arabidopsis thaliana RAD52* genes encode four open reading frames (ORFs) through differential splicing, each of which specifically localized to the nucleus, mitochondria, or chloroplast. The *A. thaliana RAD52-1A* ORF provided partial complementation to the yeast *rad52* mutant. *A. thaliana* mutants and RNA interference lines defective in the expression of *RAD52-1* or *RAD52-2* showed reduced fertility, sensitivity to mitomycin C, and decreased levels of intrachromosomal recombination compared with the wild type. In summary, computational and experimental analyses provide clear evidence for the presence of functional RAD52 DNA-repair homologs in plants.

INTRODUCTION

Genetic studies in yeast have shown that RADiation sensitive52 (RAD52) plays an essential role in homology-dependent DNA double strand break (DSB) repair and recombination (Symington, 2002). *rad52* mutants are extremely sensitive to ionizing radiation, exhibit phenotypic changes in somatic recombination pathways, including single-strand annealing, break-induced repair, and pathways involving Holliday junction intermediates, and fail to undergo meiotic recombination (Symington, 2002). Similarly, gene targeting is completely abolished in *RAD52* deletion mutants in the yeast *Kluyveromyces lactis* (Milne and Weaver, 1993). Genetically, *RAD52* forms an epistatic group with *RAD51* and *RAD54*, all of which affect the response to DNA damaging agents (Symington, 2002). RAD52 is primarily involved in repair of DNA DSBs and is also implicated in repair of other types of DNA lesions, such as stalled replication forks (Gangavarapu et al., 2007). Mechanistically, RAD52 is recruited to the Replication Protein A (RPA)-single-stranded DNA nucleoprotein complex, formed upon DSB induction and exonucleolytic ends resection, and mediates its replacement by RAD51. RAD51 then catalyzes strand invasion and D-loop formation. Eventually, RAD52 may assist in capturing the second DNA end and promote its annealing to the D-loop, thus leading to the formation of a Holliday junction (Mortensen et al., 2009).

The RAD52 protein has several domains, each with a distinct function. Its N-terminal is the most conserved domain across eukaryotic homologs (Park et al., 1996; Shen et al., 1996; Shinohara et al., 1997; Hays et al., 1998; Mer et al., 2000; Stasiak et al., 2000; Ranatunga et al., 2001). This domain enables the formation of a ring-shaped oligomer by including sites for self-association and in yeast for interaction with its RAD59 paralog (Shinohara et al., 1998; Stasiak et al., 2000; Ranatunga et al., 2001). The external part of the ring formed in the human homolog was shown to include a positively charged, DNA binding groove (Kagawa et al., 2002; Singleton et al., 2002). The central portion contains a domain for interaction with RPA (Hays et al., 1998). The C terminus, which interacts with RAD51 and also binds single- or double-stranded DNA (Milne and Weaver, 1993; Mortensen et al., 1996), demonstrated the lowest degree of conservation across species.

Yeast and animal RAD52 proteins are similar in sequence and function, but also have some significant variances. Although disruption of the mouse *RAD52* homolog strongly reduces gene targeting, it does not cause sterility or sensitivity to ionizing radiation (Rijkers et al., 1998), suggesting either a less critical role when compared with its yeast homolog or genetic redundancy. Genetic analysis of single and double mutant cell lines suggests that the human *RAD51* paralog *XRCC3* may partially compensate for the absence of *RAD52* (Fujimori et al., 2001). Similarly, BRCA2 demonstrates functional similarity to RAD52, in that it interacts with RAD51 and loads it on an RPA-covered nucleofilament (San Filippo et al., 2006; San Filippo et al., 2008). In addition, numerous mammalian RAD52 isoforms, generated by

[1] These authors contributed equally to this work.
[2] Address correspondence to avi.levy@weizmann.ac.il.
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Shmuel Pietrokovski (shmuel.pietrokovski@weizmann.ac.il) for sequence and structural data and Avraham A. Levy (avi.levy@weizmann.ac.il), for biological materials and constructs.
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.111.091744

alternative splicing, code for protein variants that may have distinct roles during homologous recombination (Thorpe et al., 2006). In chicken cells, *RAD52* is required for homologous recombination but differs from its yeast homolog in other DNA repair functions (Sonoda et al., 2001).

In plants, *RAD51* (Doutriaux et al., 1998; Bleuyard et al., 2006) and *RAD54* (Osakabe et al., 2006; Shaked et al., 2006) homologs have been identified, leaving the question of whether plants contain *RAD52* homologs (Iyer et al., 2002; Mortensen et al., 2009) or whether their absence may be compensated for by two plant *BRCA2* homologs (Siaud et al., 2004; Abe et al., 2009). Nevertheless, the absence of *RAD52* in plants is somewhat intriguing, considering its role in mediating critical steps of homologous recombination and its relatively high conservation across eukaryotes. This article describes the search for, discovery, and characterization of plant *RAD52* homologs. Computational and experimental evidence confirming the correspondence of the plant homologs to yeast and animal RAD52 proteins is provided. In addition, functional analysis of *Arabidopsis thaliana* RAD52 homologs and localization of their splice variant isoforms was performed. These data collectively confirm the presence of *RAD52* DNA-repair homologs in plants.

RESULTS

Higher Plants Include a Family of *RAD52* Homologs Encoding Several Genes and Splicing Isoforms

Plant *RAD52* homologs were first identified in the *A. thaliana* genome by querying it with the yeast RAD52 protein using the PSI BLAST program. A putative protein was detected, (At1g71310, accession number NP_849876.1), with an E-value of $1e^{-04}$ and 31% amino acid sequence identity to the yeast RAD52 56-176 N-terminal residues. Next, multiple sequence alignments (MSAs) of land plant homologs of the identified *A. thaliana* RAD52-like protein were constructed (see Supplemental Data Sets 1 and 2 online). These plant proteins are 170 to 220 residues long, with weakly conserved N termini and a well-conserved 129-residue-long region at the central and C-terminal segments. This latter region was found to be significantly similar, with an E-value of 7.8×10^{-6}, (Compass program [Sadreyev and Grishin, 2003]) to MSAs of RAD52 proteins from animals, fungi, and lower eukaryotes in the RAD52 catalytic domains for homologous pairing and multimerization (Figure 1; see Supplemental Data Set 3 online). Of note, human residues found to play an important role in DNA binding (magenta circles in Figure 1) tend to be conserved across kingdoms, including in plants.

Plant RAD52 homologs were identified by thorough sequence analysis of various available plant data sources covering proteins, expressed sequence tags (ESTs), transcriptomes, and genome data. Full and partial protein sequences were assembled from diverse land plants, ranging from liverworts and mosses to flowering plants, and from charophyte unicellular green algae (Figure 2). The plant RAD52 homolog family was categorized into two subtypes, each of whose members were found in gymnosperms, monocots, dicots, and liverworts. Some lineages have only one homolog type (e.g., mosses with only type 2 and charophytes with only type 1) (Figure 2). The sequence similarity between the two RAD52 homolog types was typically 40 to 67% across the C-terminal two-thirds (120 to 150 amino acids) of the protein sequence (see Supplemental Figure 1 online). The presence of the two homolog types suggests a gene duplication that occurred as early as the appearance of nonvascular land plants. Later duplications were apparent in specific lineages. These duplicates are separate genes within the same strain, rather than different alleles in the same strain or in different strains of the same species. This was verified by finding the duplicates in different genomic contexts, by finding several duplicates in single strains, and by finding common duplicates in several related species, as in the case of the grasses, which include two type-2 RAD52 homologs. Mosses, represented by two very well-sequenced species (*Physcomitrella patens* [Rensing et al., 2008] and *Syntrichia ruralis* [Oliver et al., 2004]), seem to include only type-2 RAD52 homologs. *P. patens* has two type-2 RAD52 homologs that apparently underwent duplication within the mosses lineage. Although almost all duplications were in type-2 RAD52 homologs, maize (*Zea mays*) features two type-1 paralogs on chromosomes 3 and 8. Sequences of type-1 *RAD52* homologs from transcribed genes were found in the *Chara vulgaris* and *Spirogyra pratensis* charophyte unicellular green algae species (National Center for Biotechnology Information [NCBI] Sequence Read Archive database entry SRX041525; Timme and Delwiche, 2010). Lack of identifiable type-2 RAD52 homologs in the charophytes' available expressed gene data could be attributable to their absence from this lineage or to specific or low expression of their genes. In either case, the type-1 and type-2 duplication apparently predates the divergence of the charophyte lineage (Figure 2). No RAD52 homologs were found in extensive data available for other major lineages of algae, including chlorophyte green algae, rhodophyte red algae, haptophyte algae, and heterokont (stramenopiles, including brown algae, diatoms, and oomycetes).

The described sequence similarity allowed us to model the structure of plant RAD52 homologs on the determined structures of the N-terminal domain of human RAD52 (Kagawa et al., 2002; Singleton et al., 2002). As expected by the relatively long corresponding regions and few insertion/deletion points (Figure 1; see Supplemental Figure 1 online), the same topology and structure features of human RAD52 are present in the models of *A. thaliana* RAD52 homologs (see Supplemental Figure 2 online). The residues corresponding to the two known RAD52 DNA binding sites (Kagawa et al., 2002; Singleton et al., 2002) are mostly conserved in the plant RAD52 homologs (Figure 1) and form similar sites, including both the positively charged groove and the second DNA binding site (see Supplemental Figures 2B and 2C online). The alpha helix of the RAD52 stem region (residues 145 to 163 in At1g71310, corresponding to residues 159 to 134 in human RAD52) demonstrated a repetitive pattern of sequence conservation. Residues facing and interacting with the stem region beta sheet were highly conserved in the plant RAD52 homologs, whereas residues facing the predicted single-stranded DNA binding site were not conserved (Figure 1; see Supplemental Figure 2D online). These features reinforce the functional homology between the plant RAD52 homolog family and the known RAD52 proteins, suggest similar function for

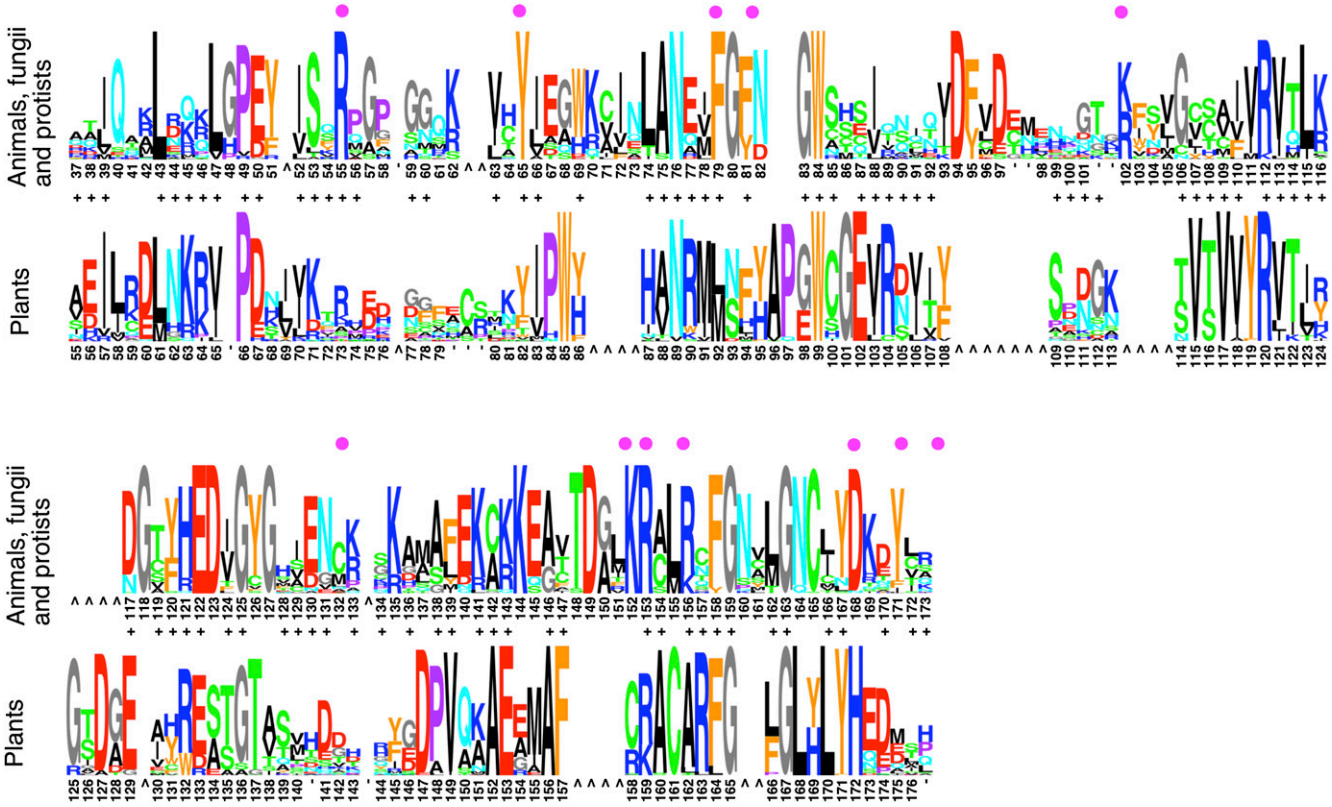

Figure 1. Sequence Conservation of Plant RAD52 Protein Family and Its Similarity to Other RAD52 Proteins.

Multiple alignments of the plant RAD52 protein family and of the animal, fungi, and protist RAD52 protein families are displayed as sequence logos. + indicates positions with similar amino acids compositions. The plant RAD52 family alignment positions are numbered according to *A. thaliana* Rad52-1B (At1g71310.1 and At1g71310.2), and the animal, fungi, and protist RAD52 family alignment positions are numbered according to human RAD52. Positions marked by dashes indicate insertions relative to the numbering protein (e.g., one or more plant RAD52 proteins have an amino acid insertion between the positions corresponding to *Arabidopsis* Rad52-1B positions 65 and 66). ^ indicates gaps introduced to align the two families. Magenta circles mark known DNA binding site residues in human RAD52.

these two families, and should assist molecular studies of plant RAD52 proteins.

RAD52-1 (At1g71310) and *RAD52-2* (At5g47870), the two identified *A. thaliana RAD52* homologs, included three splice variants for *RAD52-1*, encoding open reading frames (ORFs) *RAD52-1A* and *RAD52-1B*. Two splice variants were found for *RAD52-2*, encoding ORFs *RAD52-2A* and *RAD52-2B* (Figure 3). The At1g71310.3 (*RAD52-1A* ORF) splice variant bore three exons, whereas splice variants At1g71310.1 and At1g71310.2 both encoded the same ORF (*RAD52-1B*), and each features four exons but differs in the number of introns and in the length of the 3′ untranslated region (UTR). The *RAD52-2A* transcript had at least two exons, and the *RAD52-2B* transcript had three exons (Figure 3).

A. thaliana RAD52s Localize in the Nucleus, Mitochondria, and Chloroplasts

Each of the *A. thaliana RAD52* ORFs was fused to enhanced green fluorescent protein (EGFP), and the resulting constructs were transiently expressed in *A. thaliana* seedlings, roots, and cell culture protoplasts. Protein localization was determined in cells using confocal microscopy to detect colocalization of the EGFP fluorescence and cellular markers (Figure 4). RAD52-1A was found, in the three cell types (Figure 4A; see Supplemental Figures 3A and 3B online), to colocalize with VirE2-nuclear localization signal (NLS)-mRFP and 4′,6-diamidino-2-phenylindole-positive controls for nuclear localization. RAD52-1A was homogenously expressed throughout the nucleus, with the exception of the nucleolus. Note that EGFP alone does not localize to the nucleus (see Supplemental Figure 3C online). RAD52-1B colocalized with the mitochondrial Sc-COX4-Mito-mCherry (Figure 4B). Unlike RAD52-1A, RAD52-2A expression was only apparent in foci at the periphery of the nucleus and chloroplast (Figure 4C; see Supplemental Figure 3D online). However, in tobacco (*Nicotiana tabacum*) BY2 cells, RAD52-2A was evenly distributed throughout the nucleus (see Supplemental Figure 3E online) RAD52-2B demonstrated a punctate expression pattern within the chloroplast (Figure 4D).

The differential localization between RAD52-1A and RAD52-1B was hypothesized to relate to their different C-terminal regions,

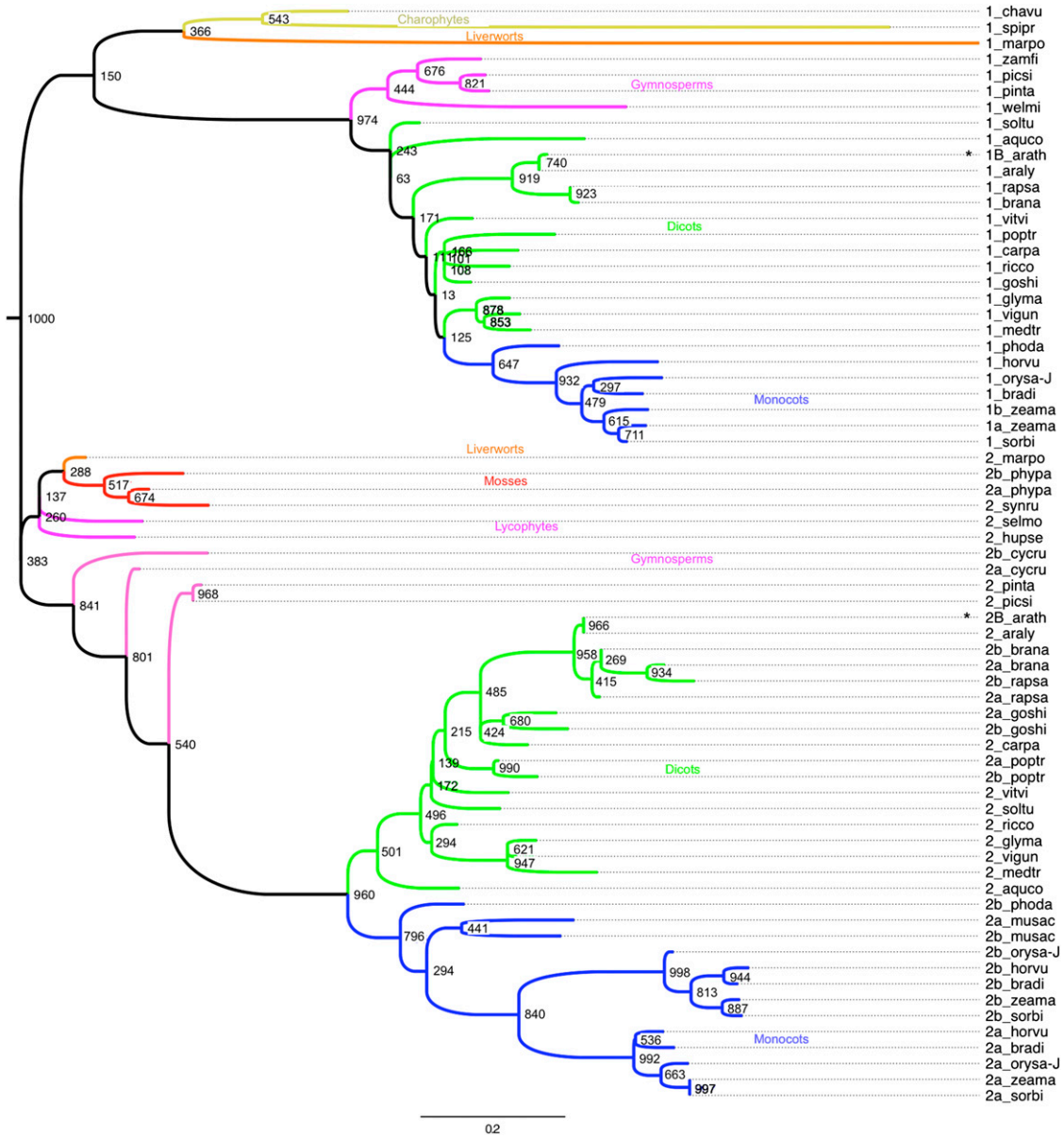

Figure 2. Plant RAD52 Homolog Phylogeny.

A dendogram of plant RAD52 homologs was calculated from the MSA shown in Figure 1 and Supplemental Data Set 2 online. The dendogram was outgroup-rooted from the position of animal, fungal, and protist (human, *Saccharomyces cerevisiae*, *Caligus rogercresseyi* [crustacean], *Trichoplax*, and *Entamoeba*) RAD52 sequences that were added to the alignment. All the added species clustered together on the dendogram. Bootstrap support values from 1000 replicates are shown in each node. The scale bar shows the number of amino acid substitutions per branch length. Branches are colored by major systematic divisions and labeled accordingly. Prefixes "1" or "2" denote the RAD52 subgroup type, and "a" or "b" indicate that more than a single gene of a particular homolog type was present in a given species. *A. thaliana* proteins are marked by asterisks. chavu, *C. vulgaris*; spipr, *S. pratensis*; marpo, *M. polymorpha*; zamfi, *Zamia fischeri*; picsi, *Picea sitchensis*; pinta, *Pinus taeda*; welmi, *Welwitschia mirabilis*; soltu, *Solanum tuberosum*; aquco, *Aquilegia coerulea*; arath, *A. thaliana*; araly, *A. lyrata*; rapsa, *Raphanus sativus*; brana, *Brassica napus*; vitvi, *Vitis vinifera*; poptr, *Populus trichocarpa*; carpa, *Carica papaya*; ricco, *Ricinus communis*; goshi, *Gossypium hirsutum*; glyma, *Glycine max*; vigun, *Vigna unguiculata*; medtr, *Medicago trunculata*; phoda, *Phoenix dactylifera*; horvu, *Hordeum vulgare*; orysa-J, *Oryza sativa* japonica cultivar; bradi, *Brachypodium distachyon*; zeama, *Z. mays*; sorbi, *Sorghum bicolor*; phypa, *P. patens*; synru, *S. ruralis*; selmo, *Selaginella moellendorffii*; hupse, *Huperzia serrata*; cycru, *Cycas rumphii*; musac, *Musa acuminata*.

The dendogram was drawn using the FigTree program (http://tree.bio.ed.ac.uk/software/figtree).

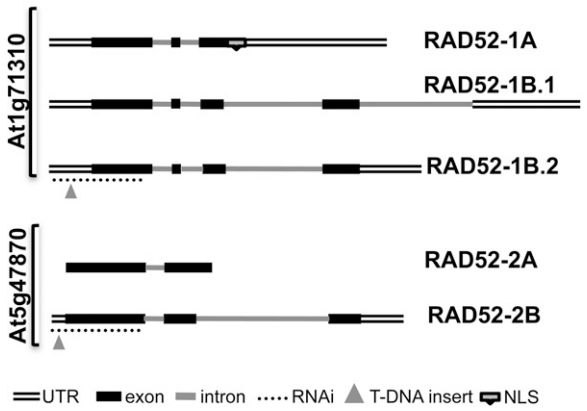

Figure 3. *A. thaliana RAD52* Gene Transcripts.

There are three known splice variants *for RAD52-1* and two for *RAD52-2*. *RAD52-1B.1* and *RAD52-1B.2* are the two *RAD52-1* (At1g71310) cDNAs, with 531-bp-long ORFs; *RAD52-1A* has a 498-bp-long ORF. T-DNA mutant SAIL_25_H08 insertion is in the 5′ UTR, 83 bp 3′ to the putative cDNA start, and 97 bp 5′ to ATG. Dashed lines indicate the location of the 385 bp RNAi for silencing of *RAD52-1* transcripts through targeting of the first exon, starting at 134 bp 5′ to ATG and ending at 251 bp 3′ to ATG. The NLS of RAD52-1A is marked by an open box. *RAD52-2* (At5g47870) cDNAs are *RAD52-2A* with a 531-bp-long ORF and *RAD52-2B* with a 600-bp-long ORF. T-DNA mutant WiscDsLox303H06 insertion is in the 5′ UTR 25 bp 3′ to the putative cDNA start and 36 bp 5′ to ATG. Dashed lines indicate the location of the 383 bp RNAi for silencing of *RAD52-2* transcripts through targeting of the first exon, starting at 61 bp 5′ to ATG, and ending at 322 bp 3′ to ATG. Full transcript lengths are 1435 bp, 2280 bp, 1559 bp, and 1502 bp for *RAD52-1A*, *RAD52-1B.1*, *RAD52-1B.2*, and *RAD52-2B*, respectively. Drawings of introns, exons, and UTRs are approximately to scale.

generated by alternative splicing (Figure 3). To test this proposal, the unique 36 amino acids C-terminal region of RAD52-1-A (VFPFLLFQCYGLIWLAFSSSRLLVEFFAFLPQKLQI), herein 52-1A-EX3-specific, was fused to EGFP and was found to colocalize with the nuclear VirE2-NLS-mRFP (Figure 4E). The 52-1A-EX3-specific sequence lacks resemblance to known NLSs, and therefore contains a novel sequence that directs nuclear localization in *A. thaliana*. In summary, RAD52 homologs are localized in all plant DNA-containing compartments. Localization of the RAD52 proteins could be predicted for mitochondrial and chloroplast compartments using the TargetP1.1 program (Emanuelsson et al., 2000) but not for the nucleus (see Supplemental Table 1 online).

Expression of *RAD52-1* and *RAD52-2* in the Wild Type, Mutant, and RNA interference *A. thaliana* Lines

RAD52 expression was analyzed using real-time PCR with splice-variant-specific primers sets (see Supplemental Figure 4 online). Expression was compared between cauline leaves, flower buds, open flowers, 4-d-old seedling roots, 4-d-old seedling shoots, ~3-mm siliques, and ~6-mm siliques. *RAD52-1A* transcription levels were similar in all these tissues, with ~1.5-fold higher expression in seedling roots (see Supplemental Figure 4A online). *RAD52* transcripts, except *RAD52-1A*, showed a marked increase of ~100-fold in plant tissues of 4-d-old seedling shoots and roots (see Supplemental Figures 4B to 4E online) when compared with other tested tissues. Expression of each splice variant was also compared in 16-d-old whole wild-type seedlings grown in the presence or absence of 10 μg/mL mitomycin C (MMC) (see Supplemental Figure 4F online). Treatment with MMC was 6 d, from day 9 to day 16. The changes we observed at this stage were relatively minor, even if in some cases statistically significant.

Mutant and RNA interference (RNAi) lines with altered *RAD52-1* and *RAD52-2* expression were generated and characterized to study the functional roles of these genes. A homozygous line was prepared from a heterozygote line containing the *rad52-1* SAIL_25_H08 T-DNA insertion allele. The homozygote *rad52-1* progeny were obtained in one-quarter of the plants, as expected for Mendelian inheritance. Homozygous mutants were also prepared for the *rad52-2* WiscDsLox303H06 T-DNA insertion allele, giving rise to *rad52-2*. Only two *rad52-2* homozygotes were found among the 46 progeny obtained from heterozygote *RAD52/rad52-2* plants, a ratio significantly lower than that expected via Mendelian inheritance (P [χ^2] = 0.0012). Such deviation from classical Mendelian inheritance could be attributable to various causes, including partial lethality of the *rad52-2/rad52-2* homozygote embryo or partial lethality of gametes (males or females or both) containing the recessive *rad52-2* allele. Embryo lethality is expected to yield a 1 (wild type): 2 (heterozygote *RAD52/rad52-2*) ratio in the progeny of heterozygous plants. Out of the 46 progeny of *RAD52/rad52-2* plants, 15 were genotyped as the wild type, whereas 29 were genotyped as heterozygous. This is in accordance (P [χ^2] = 0.9151) with the 1:2 ratio, supporting that the *rad52-1* gametes were viable and that embryo viability was compromised. Note that a 1 (wild type): 1 (heterozygote) model that would be caused by gamete lethality in male or female lineage was rejected (P [χ^2] = 0.034), further indicating that there is no major problem of gamete viability. The location of the T-DNA insertion within each gene is depicted in Figure 3 and described at the sequence level (see Methods and Supplemental Figure 5 online).

RNAi lines were constructed for repression of *RAD52-1* and *RAD52-2* transcripts (see Methods). Three single-locus T-DNA lines of each RNAi construct were selected for further analyses. *RAD52* expression was determined by real-time PCR on mRNA isolated from seedlings of the wild type, mutant, and RNAi lines (Figure 5). Expression analysis of the *RAD52-1* transcripts in the various lines showed that *rad52-1* exhibited almost no expression compared with the wild type; *RAD52-1* RNAi lines (52-1 RNAi) showed reduced expression ranging from moderate to ~60,000-fold less than the wild type (Figure 5, line 52-1 RNAi3). *rad52-1* and line 52-1 RNAi3 were selected for further characterization of the effect of reduced *RAD52-1* expression. Upon RNAi inhibition of *rad52-2*, *RAD52-1* exhibited slightly higher transcript levels than the wild type. This was not statistically significant but was consistent in all RNAi lines, suggesting the possibility of compensation. Similarly, expression analysis of the *RAD52-2* transcript in the various lines showed that the *rad52-1* and *RAD52-1* RNAi lines had slightly more *RAD52-2* transcript than the wild type. The expression of the *RAD52-2* transcript in

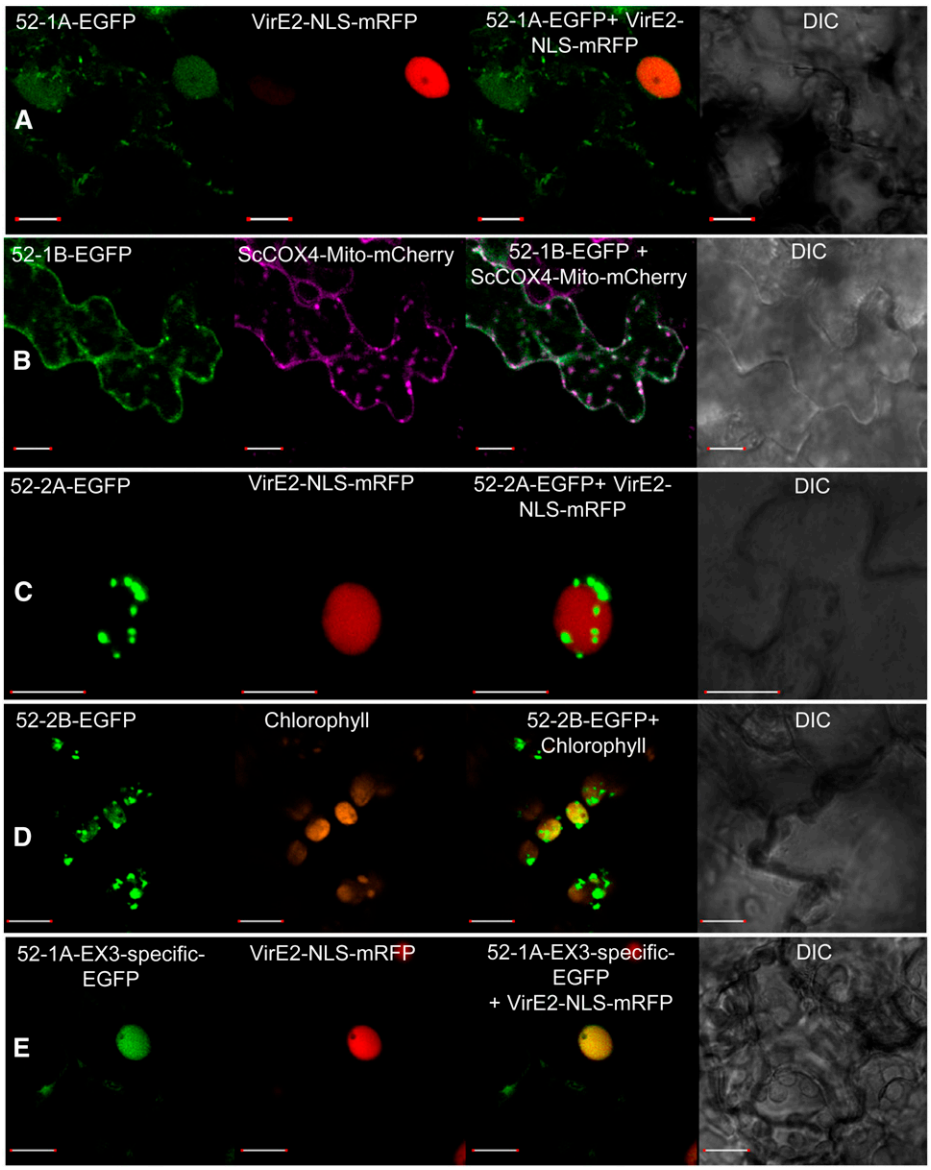

Figure 4. Cellular Localization of RAD52 Proteins in *A. thaliana* Seedlings.

Four-day-old *A. thaliana* seedlings were transformed with agrobacteria carrying the fusion protein constructs (A) 52-1A-EGFP and VirE2-NLS-mRFP; (B) 52-1B-EGFP and Sc-COX4-Mito-mCherry; (C) 52-2A-EGFP and VirE2-NLS-mRFP; (D) 52-2B-EGFP; and (E) 52-1A-EX3-specific and VirE2-NLS-mRFP. VirE2-NLS-mRFP is a nuclear marker. Sc-COX4-Mito-mCherry is a mitochondrial marker. 52-1A-EX3-specific is a short sequence that is unique to RAD52-1A. In each panel, the first picture is of EGFP, the second is of the localization marker, the third is an overlay of the first two, and the last is an image captured under visible light (differential interference contrast [DIC]).
Bars = 10 μm.

the *rad52-2* mutant was reduced by approximately twofold. *RAD52-2* RNAi (52-2 RNAi) lines showed a reduction in expression ranging from fivefold to 12-fold compared with the wild type (Figure 5). *rad52-2* and line 52-2 RNAi8 were selected for further characterization of the *RAD52-2* reduced expression phenotype. The *rad52-1* mutant and 52-2 RNAi lines were combined for a most effective double knockdown. Expression of *RAD52-1* and *RAD52-2* in the double knockdown plants was similar to the single knockdown lines from which they originated, *rad52-1* and 52-2 RNAi8 (Figure 5).

It was recently discovered that the telomerase RNA1 (*TER1*) transcript overlaps with the first two exons and introns of *RAD52-1* (Cifuentes-Rojas et al., 2011). *TER1* is the RNA subunit of telomerase and serves as a template for telomeric DNA addition. *TER1* transcript levels and the consequential telomerase activity could potentially be affected by the *rad52-1* mutant

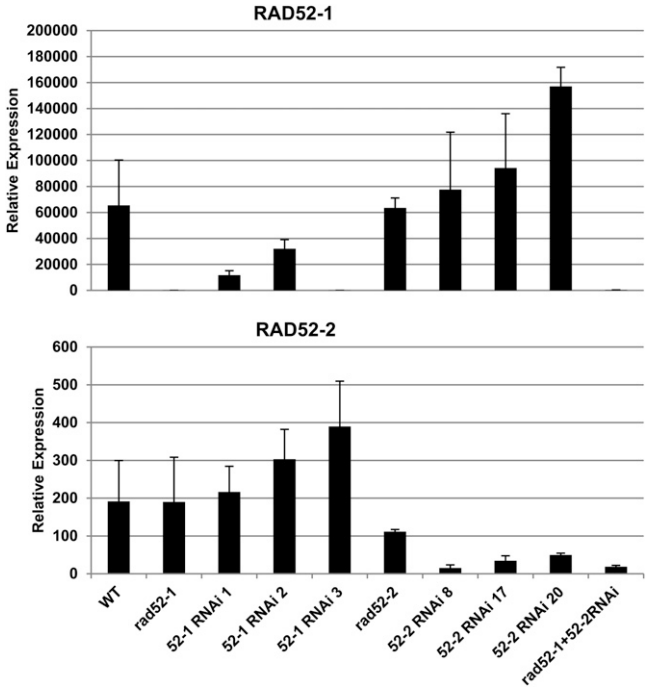

Figure 5. *RAD52* RNA Expression in Knockout Lines.

RNA was extracted from 8-d-old seedlings. Expression levels of *RAD52-1* **(Top)** and *RAD52-2* **(Bottom)** were determined by real-time PCR and normalized relative to ubiquitin expression. *RAD52-1* primers were selected from a region common to all splice variants. *RAD52-2* primers were selected from a region specific to the *RAD52-2* B splice variant (see Methods). Values represent the average of three independent repeats of the experiment. Error bars represent SE. WT, wild type.

and the *RAD52-1* RNAi and might influence the phenotype of these lines. For this reason, *TER1* transcript levels were analyzed in inflorescences, where *TER1* is highly expressed. A sixfold and threefold decrease in *TER1* expression was observed in the *rad52-1* mutant and the *RAD52-1* RNAi line, respectively, when compared with the wild type (see Supplemental Figure 6 online). However, terminal restriction fragment (TRF) analysis was also performed, and no decrease in telomere length was detected in either line when compared with the wild type (see Supplemental Figure 6 online).

Reduced Fertility of *rad52-1* and *rad52-2* Mutants and RNAi Lines

Significant reductions in fertility, expressed as the number of seeds per silique, correlated with reduced expression of *RAD52* genes in the tested mutant and RNAi lines (Figure 6). However, no additive reduction in fertility was observed upon combination of *rad52-1* and *RAD52-2* RNAi when compared with each of these lines separately. The measured reduction in seed number per silique in the various knockdown lines could be caused by the involvement of *RAD52-1* and *RAD52-2* in the viability of either or both somatic and meiotic cell types.

Role of *RAD52-1* and *RAD52-2* in DNA Damage Response and Homologous Recombination

A series of assays were performed to evaluate the expected role of *RAD52-1* and *RAD52-2* in plant DNA recombination and DNA damage response. In complementation assays, *RAD52-1A* complemented the yeast *rad52* methyl methanesulfonate (MMS)-sensitive phenotype to a lesser extent than the yeast *RAD52* (see Supplemental Figure 7 online). Complementation was only partial, but reproducible, suggesting that the *A. thaliana* RAD52 retained DNA repairing roles similar to those performed by the yeast RAD52 homolog.

The involvement of *A. thaliana* RAD52 in DNA damage repair was then tested by monitoring seedling growth after MMC-induced DNA damage. Interstrand cross-links mediated by MMC can lead to DSBs (Rink et al., 1996), which in turn can be repaired by homologous recombination. Both the *rad52-1* mutant and *RAD52-1* RNAi lines showed a significant reduction in seedling growth after MMC treatment (Figure 7A) compared with the wild type, suggesting defective repair capacity. Similar reductions in growth rates were also observed in the *rad52-2* mutant and *RAD52-2* RNAi lines (Figure 7B). Sensitivity to MMC treatment was not additive in the *rad52-1* and *RAD52-2* RNAi combination (Figure 7C), in agreement with the observed combination fertility phenotype.

Somatic recombination rates were assessed in the *rad52-1* mutant and *RAD52-2* RNAi lines using an intrachromosomal recombination (ICR) assay with a reporter transgene consisting of two overlapping fragments of the β-glucuronidase (GUS) (uidA) gene separated by a hygromycin selectable marker (Swoboda et al., 1994). *rad52-1* mutant plants and GUS recombination reporter plants were crossed, and the frequency of recombination events was monitored in F3 progeny plants homozygous for both the recombination reporter and the *rad52-1* mutation (see Methods). The *RAD52-2* RNAi construct was

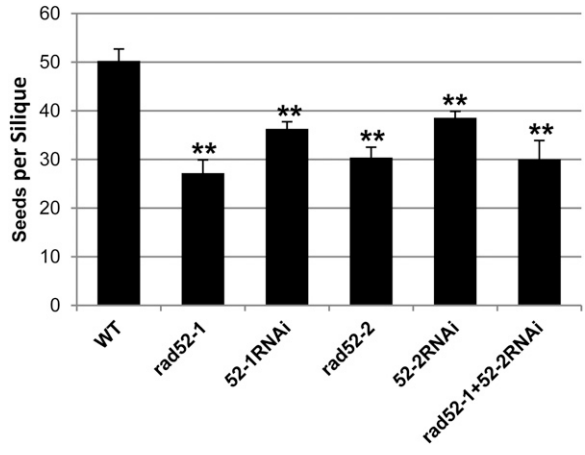

Figure 6. Reduced Fertility in *RAD52* Knockout Lines.

Plants of each genotype were grown in soil (n = 8 to 18). Seeds of each plant were collected, scored, and divided by the number of siliques per plant. Error bars represent SE. ** indicates $P < 0.01$ using analysis of variance to compare the wild type (WT) with each genotype.

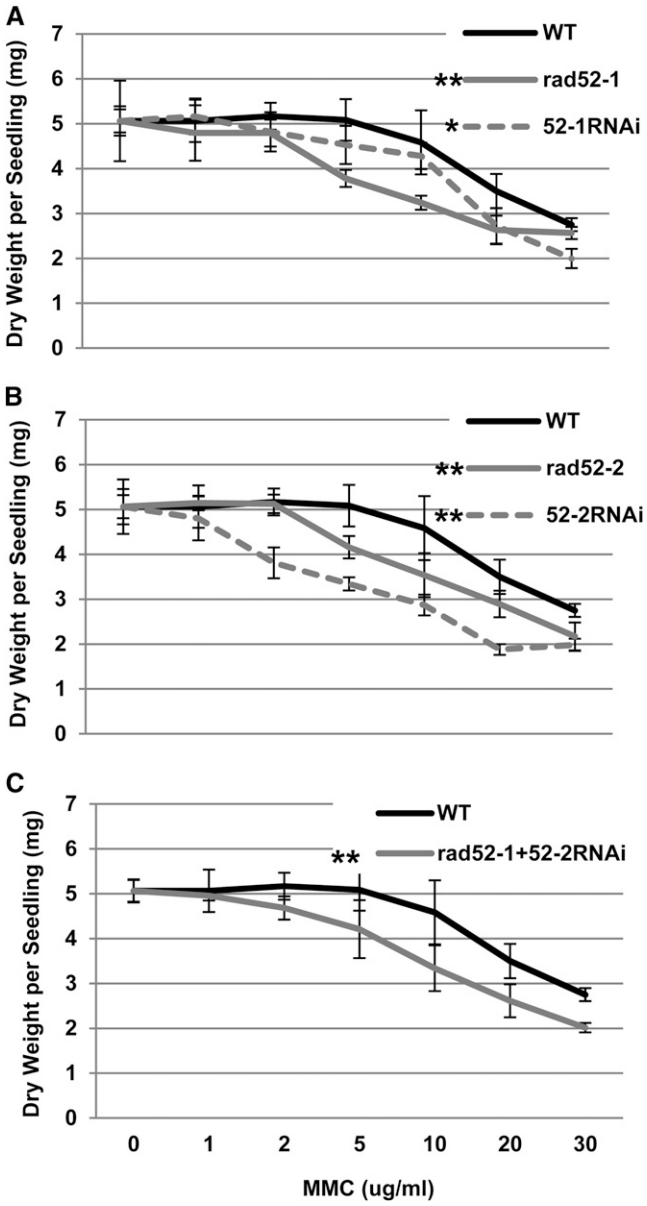

Figure 7. Sensitivity to MMC in Plants with Reduced *RAD52* Gene Expression.

For each genotype, dry weight per seedling (mg) was compared in the absence of MMC or at 6 d after treatment with 1, 2, 5, 10, 20, 30 μg/mL MMC. The wild type (WT) was compared with each genotype using two-way analysis of variance. * indicates $P < 0.05$, ** indicates $P < 0.01$. Each data point is an average of three to four repeat experiments measuring five seedlings each. Error bars represent SE.

directly transformed into the ICR line, and a line demonstrating an approximately fivefold reduction in *RAD52-2* RNA was selected for further analysis in T2 progeny plants. Somatic ICR frequencies were determined in the wild type versus mutant lines under normal or MMC-induced conditions. No difference in ICR rates was detected in the *rad52-1* mutant under normal conditions (Figure 8A); however, when seedlings were treated with MMC, ICR rates were significantly reduced in the mutant compared with the wild type (Figure 8B). In the *RAD52-2* RNAi line, a significant reduction in ICR rates compared with the wild type was detected under both normal and MMC treatment conditions (Figures 8C and 8D, respectively). Both the somatic homologous recombination phenotype and the MMC sensitivity of the *RAD52* mutants and RNAi lines are indicative of defective repair pathways.

DISCUSSION

The findings reported herein describe *RAD52* homologs in plants, detected using sensitive computational analyses of conserved domain MSAs. Moreover, conservation of predicted structures and of amino acid residues important for DNA binding of human RAD52 (Kagawa et al., 2002) (Figure 1; see Supplemental Figure 1 online) strongly suggest a common function between plant RAD52 and the N-terminal domains of previously described RAD52 proteins.

The long forms of the RAD52 proteins in fungi and animals demonstrate high sequence similarity in their N-terminal domains. By contrast, the C-terminal segments of these RAD52 proteins contain the RAD51- and RPA-interaction regions and exhibit weak similarity (Park et al., 1996; Shen et al., 1996; Hays et al., 1998; Shinohara et al., 1998; Mer et al., 2000; Stasiak et al., 2000; Ranatunga et al., 2001). Moreover, the short forms of RAD52, resulting from alternative splicing, have been found to mostly contain the conserved N-terminal domain (Kito et al., 1999). Yeasts also include RAD59, which is homologous to the N-terminal domain of RAD52 and has overlapping functions with RAD52 (Kagawa et al., 2002). Bacteriophage homologs of RAD52 (Iyer et al., 2002) also only bear the RAD52 N-terminal domain (Ploquin et al., 2008). The plant RAD52 protein sequences resemble the short form that has only the N-terminal domain (Figure 1).

The plant RAD52 family can be divided in two subfamilies, unlike animals and several protists. Therefore, we conclude that an early gene duplication within the family occurred at the emergence of nonvascular land plants (Figure 2). The separation of the two plant RAD52 types is robust (bootstrap support value of 1000/1000) and apparent in the nonvascular land plant *Marchantia polymorpha* (a liverwort), but the positions of nonseed plant subgroups within either subgroup is less certain (Figure 2). Unlike the liverwort homologs, other nonseed plant representatives bear either one or the other homolog type. Mosses and lycophytes, whose sequenced genome is of high quality and fairly complete, feature the type-2 RAD52 homologs only (Figure 2). These nonseed plants probably lost the plant *RAD52* type-1 gene presumably present in their progenitors. Presence of the *RAD52* type-1 gene in two charophyte species indicates a common algal ancestor to all plant *RAD52* genes, because land plants emerged from the charophyte green algae lineage (Finet et al., 2010). RAD52 homologs were not detected in chlorophytes, the other major green algae lineage, or other major lineages of algae (including red algae and diatoms). This divergence suggests that plant *RAD52* genes serve as molecular markers of land plants and their specific green algae predecessors. In addition, a question

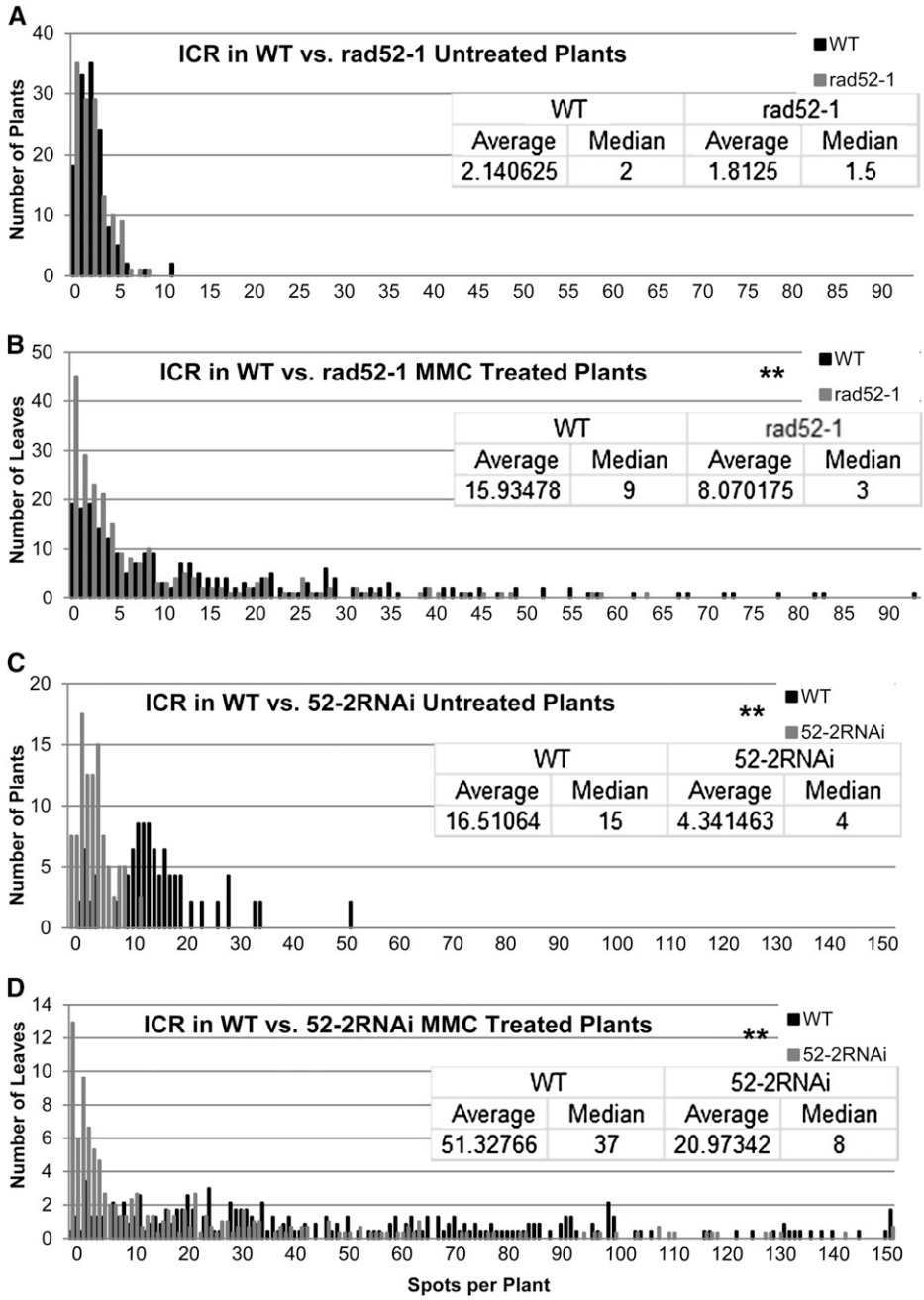

Figure 8. DNA Damage Response and Homologous Recombination in Plants with Reduced *RAD52* Gene Expression.

Somatic recombination rates were evaluated using the GUS ICR. Results are presented per plant for untreated plants [**(A)** and **(B)** $n = 128$ for each line; **(C)** and **(D)** $n = 47$ wild type (WT) and 40 52-2 RNAi] and per leaf for 2 μg/mL MMC-treated plants [**(A)** and **(B)** $n = 230$ for each line; **(C)** and **(D)** $n = 235$ wild type and 302 52-2 RNAi), because of higher recombination rates. The wild type was compared with each genotype using the Wilcoxon nonparametric test. ** indicates $P < 0.01$.

arises whether other genes substitute for the absence of *RAD52* homologs in other algae.

In addition to gene duplication, alternative splicing contributes to the diversity of *A. thaliana* RAD52 proteins and allows for their localization within all DNA-containing cellular compartments (Figures 3 and 4). The consistent sublocalization of the various RAD52 proteins within these organelles suggests their putative role in maintaining nuclear and organellar genomes. RAD52-1A is localized within the nucleus. This splice variant contains a unique C-terminal 36 amino acid region (52-1A-ex3-specific) that

promotes EGFP accumulation in the nucleus and represents a new NLS sequence in *A. thaliana*. Unlike RAD52-1A, which was localized uniformly throughout the nucleus, RAD52-2A showed foci that were confined to the periphery of the nucleus, possibly within or outside of the nuclear envelope. To further investigate the atypical nuclear localization of RAD52-2A, we also tested localization in tobacco BY2 cells. *A. thaliana* RAD52-2A was uniformly distributed in tobacco nuclei (see Supplemental Figure 3E online). Although these experiments confirm the nuclear association of RAD52-2A, the discrepancy between the *A. thaliana* and tobacco localization data needs to be further studied. Mitochondrial localization of RAD52-1B hints at a functional role similar to that of the yeast mitochondrial genome maintenance101 (MGM101). The mitochondrial MGM101 is required for mitochondrial DNA replication or transmission. MGM101 features an active core with some degree of structural similarity with the RAD52 protein family. MGM101 also contains only the RAD52 N-terminal domain (Zuo et al., 2007). The DNA-maintenance machinery is also required for normal chloroplast function. For example, chloroplast DNA DSB repair by homologous recombination was demonstrated in *Chlamydomonas* (Odom et al., 2008); and in *A. thaliana*, a chloroplast-targeted homolog of RecA encoded by the nuclear genome was shown to be involved in chloroplast DSB repair (Rowan et al., 2010). Plant RAD52 homologs targeted to the chloroplast, such as *A. thaliana* RAD52-2, could take part in the chloroplast DNA repair pathway. The dual localization of RAD52-2A in chloroplasts and the nucleus is intriguing and could be a part of the crosstalk between the nucleus and the chloroplast (reviewed in Woodson and Chory, 2008). A recent study in *A. thaliana* shows that protein cleavage of a chloroplast-bound protein releases a peptide that mediates chloroplast signals to the nucleus (Sun et al., 2011).

The functional assays described in this article provide evidence for the role of RAD52 in DNA recombination and repair and possibly in other pathways. This includes the partial complementation by *RAD52-1A* of a yeast *rad52* null mutant phenotype of sensitivity to MMS (see Supplemental Figure 7 online), which suggests a similar role of the plant protein in DNA repair. In addition, several assays point to a somatic role for the *A. thaliana* RAD52 homologs, as demonstrated by their subcellular localization (Figure 4) and by the sensitivity (Figure 7) and reduced ICR rates (Figure 8) of mutants or RNAi line seedlings and leaves, respectively, to MMC treatment.

The reduced fertility phenotype, expressed as seeds per silique, in mutants and RNAi lines (Figure 6) may be caused by defective somatic and/or meiotic DNA recombination and repair. Testing this will require in-depth analysis of meiosis in these defective plants. The possibility that reduced fertility in *rad52-1* was caused by the disruption of the overlapping *TER1* gene seems unlikely, because despite a reduction of three- to sixfold in TER1 transcript in the *rad52-1* mutant, telomere length remained unchanged (see Supplemental Figure 6 online). Moreover, proximity to a *TER* gene does not exist for both *A. thaliana RAD52* genes, but reduced fertility was still observed in *RAD52-2* knockdown lines. Furthermore, the *TER1 RAD52* gene overlap is evolutionarily unstable. The absence of *TER1* gene telomerase repeats, along with the conservation of only the *TER1* region overlapping the *RAD52-1* transcript in *Arabidopsis lyrata*, suggests that these loci are unlinked. Further supporting evidence can be taken from the lack of telomerase repeats (the hallmarks of *TER* genes) near *RAD52-1* homologs in more distant species (data not shown).

The reduced fertility in *rad52-2* may also be caused by somatic defects. Indeed, it was difficult to obtain a homozygous mutant; only two of 46 progeny were homozygotes, whereas the proportion of heterozygotes was as expected for Mendelian inheritance. This observation suggests partial lethality in the embryo; however, meiotic defects cannot be ruled out. In addition, the effect of *RAD52-2* RNAi on ICR in young, untreated seedlings was more severe compared with *rad52-1* (Figure 8). Alternatively, the partial lethality of *rad52-2* could be related to *RAD52-2* localization within the chloroplast, as described for other chloroplast-localized proteins (Budziszewski et al., 2001; Zuo et al., 2007; Chigri et al., 2009). In this case, maintenance of the stability of the chloroplast genome may be essential for metabolic chloroplast functioning during seed development.

A combination of *rad52-1* and *RAD52-2* RNAi did not decrease fertility (Figure 6) or sensitivity to MMC (Figure 7) when compared with the single lines. There are several interpretations of this finding. One possibility is that the plant RAD52, much like the human RAD52, is likely to form a multimer. The detected similarity between the two lies within a DNA binding domain, which also forms a multimer (Kagawa et al., 2002). Loss of either subunit would then cause the same phenotype. Alternatively, an unknown additional gene might compensate for the loss of one or both *A. thaliana RAD52* genes.

In summary, this article outlines the sequence, structure, and suggested functional roles of plant *RAD52* homologs. In contrast with the essential yeast *RAD52*, plant homologs bear a moderate effect on homologous recombination. In this respect, the plant *RAD52* gene demonstrates a closer resemblance to human *RAD52*. The plant *RAD52*-like family shows a complex pattern of expression and localization, with multiple genes and splice variants. Further work will be required to elucidate the mechanisms through which the diverse plant RAD52 proteins affect the DNA repair and fertility phenotypes reported here.

METHODS

Bioinformatics and Phylogenetic Analysis

Sequence searches were performed on the NCBI databases using the BLAST and PSI-BLAST programs (Altschul et al., 1997). The queried databases were of protein, genomic, and transcribed (EST) sequences from the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Joint Genome Institute (http://genome.jgi-psf.org). ESTs were assembled using the CAP3 program (Huang and Madan, 1999). Multiple alignments were found using PSI-BLAST, Block Maker (Henikoff et al., 1995), Macaw (Schuler et al., 1991), and FSA (Bradley et al., 2009) programs (see Supplemental Data Sets 2 and 4 online). Sequence logos (Schneider and Stephens, 1990) were calculated as previously described (Henikoff et al., 1995). The conservation of the alignment positions was taken from the position-specific scoring matrix of protein multiple alignments. The conservation values are the information content of each position in bits, corresponding to the total height of each column of the sequence logos and to the coloring scheme in Supplemental Figure 2D online. Multiple alignment to multiple alignment comparisons were performed using the LAMA (Frenkel-Morgenstern et al., 2005) and Compass (Sadreyev and

Grishin, 2003) programs. Structure models were calculated on the LOMETS (Wu and Zhang, 2007), Phyre (Kelley and Sternberg, 2009), and mGenThreader (McGuffin and Jones, 2003) servers. Several equally stable models, based on the known structure of the human RAD52 N-terminal domain, were found. The model used in this work was calculated on the LOMETS server by the HHsearch (Söding, 2005) program. PyMol (version 0.99rc6; DeLano Scientific) was used to examine protein structures, calculate their approximate charge distribution, map the sequence conservation on the structure, and generate figures.

Phylogenetic trees were calculated in PHYML (Guindon and Gascuel, 2003), version 2.4.4, with four substitution rate categories and other default parameters to estimate gamma shape parameter, and invariant proportion. These parameters were then used with 1000 bootstrap replicates. The dendogram was outgroup-rooted from the position of animal, fungal, and protist *RAD52* sequences that were added to the alignment (Figure 2).

Mutants

T-DNA-tagged mutants were obtained from the Arabidopsis Biological Resources Center (Alonso et al., 2003). Border cloning and sequencing for each mutant homozygous line provided the corresponding gene's sequence from both ends of the T-DNA. *rad52-1* features a left border at the 5′ and a right border at the 3′, and *rad52-2* has left borders at both ends of the insertion. The *rad52-1* mutant SAIL_25_H08 T-DNA insertion is in the 5′ UTR, 83 bp 3′ to the putative cDNA start, and 97 bp 5′ to ATG (Figure 3; see Supplemental Figure 5 online). The *rad52-2* mutant WiscDsLox303H06 T-DNA insertion is in the 5′ UTR, 25 bp 3′ to the putative cDNA start, and 36 bp 5′ to ATG.

cDNA

RAD52-1 (At1g71310) cDNA sources: At1g71310.1 (*RAD52-1B.1*) 531-bp-long ORF: NM_105800: SQ036a05F (Kazusa, Japan). At1g71310.2 (*RAD52-1B.2*) 531-bp-long ORF: NM_179545: RAFL11-06-I22 (Riken, Japan) At1g71310.3 (*RAD52-1A*) 498-bp-long ORF: NM_202394: BX816488 (Institut National de la Recherche Agronomique, France). The three publically available cDNA clones of *RAD52-1* splice variants were sequenced, and their sequences were identical to the ones published in public databases. To clone *RAD52-2A* cDNA, we used RT-PCR, with the primers (see Supplemental Table 2 online for primers sequences) AR52-2_RI_F and AR522A_Xho_R, on wild-type *Arabidopsis thaliana* Columbia 4-d-old seedling roots cDNA. To clone *RAD52-2B* cDNA, we used RT-PCR, with the primers AR52-2_RI_F and AR52-2_Xba_R, on wild-type *A. thaliana* Columbia inflorescences cDNA (sequence identical to NM_124161 ORF).

Genotoxicity Assay for Yeast Cells

The yeast MK166-52 *rad52* strain (kindly provided by Martin Kupiec (Liefshitz et al., 1995) was transformed with *A. thaliana RAD52* homologs that were cloned into pGMU10 (Iha and Tsurugi, 1998) under the inducible Gal1 promoter. The BY4742 strain was used as the wild-type control. MMS sensitivity was tested by growing yeast cells to stationary phase in SD-Leu-Ura (SD-Trp for the wild type) medium and then diluting them and growing them again to logarithmic phase in liquid SC-Ura medium supplemented with either 2% Glc or Gal and 1% raffinose. Yeast were counted and plated in serial dilutions on SC + Glc in the presence or absence of MMS (0.01%) and on SC + Gal in the presence or absence of MMS (0.01%). Plates were incubated at 30°C for 5 d.

Subcellular RAD52 Localization

Four-day-old *A. thaliana* seedlings were transformed with agrobacteria carrying fusion protein constructs using a previously described method (Li et al., 2009). *A. thaliana* roots were transformed with agrobacteria carrying fusion protein constructs using a previously described method (Gelvin, 2006). Transformation of *A. thaliana* cultured cell protoplasts was performed as previously described (Yoo et al., 2007). Tobacco (*Nicotiana tabacum*) BY2 cells at 20% cells per volume stage were mixed with 1/80 volume of agrobacteria carrying fusion protein constructs in logarithmic phase and visualized 3 d after transformation. *RAD52-1* and *RAD52-2* ORFs were fused to EGFP at the C-terminal end by replacing the stop codon with an *Nco*I site. Sc-COX4-Mito-mCherry is the mitochondrial targeting signal of yeast COX4 fused to mCherry mt-rk CD3-991 (Nelson et al., 2007). The VirE2-NLS-mRFP, consisting of four repeats of the minimal 134 bp VirE2 NLS fused to mRFP, was a kind gift from Professor Yuval Eshed (Alvarez et al., 2009). All proteins, aside from Sc-COX4-Mito-mCherry, were flanked by the CaMV35S promoter and the octopine synthase 3′ polyA signal. Sc-COX4-Mito-mCherry was flanked by a double CaMV35S promoter and the nopaline synthase 3′ polyA signal. Cellular localization was analyzed using a laser confocal microscope (Olympus IX81 FV1000 Spectral) equipped with an FV1000 UPLAPO 60×O NA: 1.35 objective lens. EGFP images were captured using Argon laser (excitation, 488 nm; emission, 500- to 545-nm intervals). mRFP and mCherry-labeled proteins were viewed using a diode laser (excitation, 559 nm; emission, 575- to 620-nm filter). Chlorophyll was detected using a diode laser (excitation, 638 nm; emission, 655- to 755-nm filter). 4′,6-diamidino-2-phenylindole was viewed under a diode laser (excitation, 405 nm; emission, 425- to 475-nm intervals). Imaging was performed in a line sequential mode.

RNAi Lines

A 385-bp fragment was used for RNAi silencing of *RAD52-1* transcripts, which targets the first exon, starting at 134 bp 5′ to ATG and ending at 251 bp 3′ to ATG. A 383-bp fragment was used for RNAi silencing of *RAD52-2* transcripts, which targets the first exon, starting at 61 bp 5′ to ATG and ending at 322 bp 3′ to ATG. RNAi regions for each gene were cloned in the pKANNIBAL vector in the order of 5′ CaMV35S promoter, sense orientation, intron, antisense orientation, and 3′ octopine synthase 3′ polyA signal (Wesley et al., 2001).

Transcript Analysis by Real-Time RT-PCR

Total RNA was extracted using Tri reagent (Molecular Research Center). Three repeat experiments of eight 8-d-old seedlings of each mutant and RNAi line were evaluated for changes in *RAD52* genes expression. cDNA was synthesized using Superscript II RNase H-Reverse Transcriptase (Invitrogen Life Technologies) and oligo (dT_{18}) primer. Reactions for quantitative real-time RT-PCR on the cDNA were performed using Applied Biosystems Power SYBR Green and run on an Applied Biosystems 7300 cycler. Reactions for each tested gene in each cDNA sample were independently repeated at least three times. Ubiquitin (At5g25760) was used as a reference for cDNA quantity. Quantification of each gene was performed using Applied Biosystems 7300 software. Relative expression of a gene in a certain sample was initially obtained by dividing the gene expression level (in arbitrary units) by the ubiquitin level (in arbitrary units). Relative expression units are shown by setting the sample with the lowest or highest expression at a value of 1. The following primers were used for analysis of *RAD52* genes expression in lines with reduced expression (Figure 5): *RAD52-1* PCR primers, 52-1CF and 52-1CR, fall within a region 3′ to *RAD52-1* RNAi and are common to all *RAD52-1* splice variants. *RAD52-2B* PCR primers, 522BF and 522BR, fall within a region 3′ to *RAD52-2* RNAi. Ubiquitin PCR primers, Ubiquitin C-F and Ubiquitin C-R. The following primers were used for analysis of specific *RAD52* splice variants in different tissues and for evaluating plant responses to MMC treatment: *RAD52-1A*, 521AUF and 521 AUR. *RAD52-1B.1*, 521B1UF and 521B1UR. *RAD52-1B.2*, 521B2UF and 521B2UR.

RAD52-2A, 522AF and 522AR. *RAD52-2B* primers were as indicated above. RNA was extracted from inflorescences for analysis of *TER1* expression in *RAD52-1* knockdown lines. cDNA was synthesized using Superscript III, as described by Cifuentes-Rojas et al., (2011). TER1UF and TER1UR primers were used for the PCR (see Supplemental Table 2 online for primers sequences).

TRF Analysis

TRF was assayed as described by Cifuentes-Rojas et al. (2011).

MMC Assay

MMC assay was performed as previously described (Hartung et al., 2007), with freshly prepared MMC solutions (Sigma-Aldrich; category no. M4287).

ICR Assay

The GUS tester line (Swoboda et al., 1994) was crossed with the *rad52-1 A. thaliana* mutant. F3 seeds from plants homozygous for both the GUS recombination substrate and for the *rad52-1* were tested. The *RAD52-2* RNAi construct was transformed into the GUS tester line. T2 seeds from a line that showed fivefold reduction of *RAD52-2* expression were tested. Seeds were plated on solid one-half Murashige and Skoog medium. At 8 d, seedlings were transferred to liquid one-half Murashige and Skoog medium. After 9 d, MMC was added to a final concentration of 2 μg/mL. Medium was added to the same volume as MMC in cases of "untreated" seedlings. At 16 d, seedlings were stained for GUS activity. The number of blue spots or sectors on each seedling was visually determined using a light microscope.

Seed Counting

Plants were grown, and seeds were collected from the beginning of maturation until the whole plant dried. Seed number was calculated based on 100 seeds weight.

Accession Numbers

Sequence data from this article can be found in the EMBL/GenBank data libraries under the following accession numbers. At1g71310.1 (*RAD52-1B* ORF): NM_105800: SQ036a05F (Kazusa, Japan), At1g71310.2 (*RAD52-1B* ORF): NM_179545: RAFL11-06-I22 (Riken, Japan), At1g71310.3 (*RAD52-1A* ORF): NM_202394: BX816488 (Institut National de la Recherche Agronomique, France). At5g47870 (*RAD52-2*). Accession numbers for sequences used in phylogenetic analysis can be found in Supplemental Data Sets 1 to 3 online.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Sequence Conservation within the Plant RAD52 Protein Family and Its Similarity to Other RAD52 Proteins.

Supplemental Figure 2. Structure Models of *A. thaliana RAD52* Homologs.

Supplemental Figure 3. Nuclear Localization of RAD52-1A and RAD52-2A in *A. thaliana* Roots, Protoplasts, and Cotyledons.

Supplemental Figure 4. Expression Analysis of *RAD52* Gene Splice Variants.

Supplemental Figure 5. *A. thaliana RAD52* T-DNA Insertion Mutant Integration Sites.

Supplemental Figure 6. Analysis of *TER1* Expression and Telomerase Activity in *AtRAD52-1* Knockdown Lines.

Supplemental Figure 7. *RAD52-1A* Partly Complements a Yeast *rad52* Null Mutant.

Supplemental Table 1. *A. thaliana* RAD52 Protein Localization Prediction Using TargetP1.1 Analysis.

Supplemental Table 2. Primers List.

Supplemental Data Set 1. Plant *RAD52* Homolog Accession Numbers.

Supplemental Data Set 2. Plant RAD52 Multiple Aligned Protein Sequences.

Supplemental Data Set 3. NCBI Protein Accession Numbers of RAD52 Proteins from Various Species Used for MSA.

Supplemental Data Set 4. RAD52 Proteins from Various Species Multiple Aligned Protein Sequences.

ACKNOWLEDGMENTS

We thank Judith Korchia for the early work with the PSI BLAST search and Vladimir Kiss for his professional and kind help with the confocal microscopy. This work was supported by a European Union research grant (RECBREED). S.P. holds a Hermann and Lilly Schilling Foundation chair. A.A.L. holds the Gilbert de Botton Chair of Plant Science.

AUTHOR CONTRIBUTIONS

A.A.L. initiated, designed, and supervised the research, supervised the data analysis, and was involved in writing the article. A.S. designed and performed the experimental research and initiated the computational research, analyzed the data, and was involved in writing the article. S.P. designed and performed the computational research, analyzed its data, and was involved in writing the article. C.M.-B. designed and performed experimental research dealing with yeast complementation experiments. N.A.-R. performed research relating to MMC and ICR experiments.

Received September 15, 2011; revised November 10, 2011; accepted December 8, 2011; published December 27, 2011.

REFERENCES

Abe, K., Osakabe, K., Ishikawa, Y., Tagiri, A., Yamanouchi, H., Takyuu, T., Yoshioka, T., Ito, T., Kobayashi, M., Shinozaki, K., Ichikawa, H., and Toki, S. (2009). Inefficient double-strand DNA break repair is associated with increased fasciation in Arabidopsis BRCA2 mutants. J. Exp. Bot. **60:** 2751–2761.

Alonso, J.M., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301:** 653–657.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Alvarez, J.P., Goldshmidt, A., Efroni, I., Bowman, J.L., and Eshed, Y. (2009). The NGATHA distal organ development genes are essential for style specification in *Arabidopsis*. Plant Cell **21:** 1373–1393.

Bleuyard, J.Y., Gallego, M.E., and White, C.I. (2006). Recent advances in understanding of the DNA double-strand break repair machinery of plants. DNA Repair (Amst.) **5:** 1–12.

Bradley, R.K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., and Pachter, L. (2009). Fast statistical alignment. PLoS Comput. Biol. **5:** e1000392.

Budziszewski, G.J., et al. (2001). Arabidopsis genes essential for seedling viability: Isolation of insertional mutants and molecular cloning. Genetics **159:** 1765–1778.

Chigri, F., Sippel, C., Kolb, M., and Vothknecht, U.C. (2009). Arabidopsis OBG-like GTPase (AtOBGL) is localized in chloroplasts and has an essential function in embryo development. Mol Plant **2:** 1373–1383.

Cifuentes-Rojas, C., Kannan, K., Tseng, L., and Shippen, D.E. (2011). Two RNA subunits and POT1a are components of Arabidopsis telomerase. Proc. Natl. Acad. Sci. USA **108:** 73–78.

Doutriaux, M.P., Couteau, F., Bergounioux, C., and White, C. (1998). Isolation and characterisation of the RAD51 and DMC1 homologs from Arabidopsis thaliana. Mol. Gen. Genet. **257:** 283–291.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:** 1005–1016.

Finet, C., Timme, R.E., Delwiche, C.F., and Marlétaz, F. (2010). Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. Curr. Biol. **20:** 2217–2222.

Frenkel-Morgenstern, M., Voet, H., and Pietrokovski, S. (2005). Enhanced statistics for local alignment of multiple alignments improves prediction of protein function and structure. Bioinformatics **21:** 2950–2956.

Fujimori, A., Tachiiri, S., Sonoda, E., Thompson, L.H., Dhar, P.K., Hiraoka, M., Takeda, S., Zhang, Y., Reth, M., and Takata, M. (2001). Rad52 partially substitutes for the Rad51 paralog XRCC3 in maintaining chromosomal integrity in vertebrate cells. EMBO J. **20:** 5513–5520.

Gangavarapu, V., Prakash, S., and Prakash, L. (2007). Requirement of RAD52 group genes for postreplication repair of UV-damaged DNA in Saccharomyces cerevisiae. Mol. Cell. Biol. **27:** 7758–7764.

Gelvin, S.B. (2006). Agrobacterium transformation of Arabidopsis thaliana roots: A quantitative assay. Methods Mol. Biol. **343:** 105–113.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:** 696–704.

Hartung, F., Suer, S., and Puchta, H. (2007). Two closely related RecQ helicases have antagonistic roles in homologous recombination and DNA repair in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA **104:** 18836–18841.

Hays, S.L., Firmenich, A.A., Massey, P., Banerjee, R., and Berg, P. (1998). Studies of the interaction between Rad52 protein and the yeast single-stranded DNA binding protein RPA. Mol. Cell. Biol. **18:** 4400–4406.

Henikoff, S., Henikoff, J.G., Alford, W.J., and Pietrokovski, S. (1995). Automated construction and graphical presentation of protein blocks from unaligned sequences. Gene **163:** GC17–GC26.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Res. **9:** 868–877.

Iha, H., and Tsurugi, K. (1998). Shuttle-vector system for Saccharomyces cerevisiae designed to produce C-terminal-Myc-tagged fusion proteins. Biotechniques **25:** 936–938.

Iyer, L.M., Koonin, E.V., and Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. BMC Genomics **3:** 8.

Kagawa, W., Kurumizaka, H., Ishitani, R., Fukai, S., Nureki, O., Shibata, T., and Yokoyama, S. (2002). Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. Mol. Cell **10:** 359–371.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: A case study using the Phyre server. Nat. Protoc. **4:** 363–371.

Kito, K., Wada, H., Yeh, E.T., and Kamitani, T. (1999). Identification of novel isoforms of human RAD52. Biochim. Biophys. Acta **1489:** 303–314.

Li, J.F., Park, E., von Arnim, A.G., and Nebenführ, A. (2009). The FAST technique: A simplified Agrobacterium-based transformation method for transient gene expression analysis in seedlings of Arabidopsis and other plant species. Plant Methods **5:** 6.

Liefshitz, B., Parket, A., Maya, R., and Kupiec, M. (1995). The role of DNA repair genes in recombination between repeated sequences in yeast. Genetics **140:** 1199–1211.

McGuffin, L.J., and Jones, D.T. (2003). Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics **19:** 874–881.

Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C.J., Edwards, A.M., and Chazin, W.J. (2000). Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. Cell **103:** 449–456.

Milne, G.T., and Weaver, D.T. (1993). Dominant negative alleles of RAD52 reveal a DNA repair/recombination complex including Rad51 and Rad52. Genes Dev. **7:** 1755–1765.

Mortensen, U.H., Lisby, M., and Rothstein, R. (2009). Rad52. Curr. Biol. **19:** R676–R677.

Mortensen, U.H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. Proc. Natl. Acad. Sci. USA **93:** 10729–10734.

Nelson, B.K., Cai, X., and Nebenführ, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J. **51:** 1126–1136.

Odom, O.W., Baek, K.H., Dani, R.N., and Herrin, D.L. (2008). Chlamydomonas chloroplasts can use short dispersed repeats and multiple pathways to repair a double-strand break in the genome. Plant J. **53:** 842–853.

Oliver, M.J., Dowd, S.E., Zaragoza, J., Mauget, S.A., and Payton, P.R. (2004). The rehydration transcriptome of the desiccation-tolerant bryophyte Tortula ruralis: Transcript classification and analysis. BMC Genomics **5:** 89.

Osakabe, K., Abe, K., Yoshioka, T., Osakabe, Y., Todoriki, S., Ichikawa, H., Hohn, B., and Toki, S. (2006). Isolation and characterization of the RAD54 gene from Arabidopsis thaliana. Plant J. **48:** 827–842.

Park, M.S., Ludwig, D.L., Stigger, E., and Lee, S.H. (1996). Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells. J. Biol. Chem. **271:** 18996–19000.

Ploquin, M., Bransi, A., Paquet, E.R., Stasiak, A.Z., Stasiak, A., Yu, X., Cieslinska, A.M., Egelman, E.H., Moineau, S., and Masson, J.Y. (2008). Functional and structural basis for a bacteriophage homolog of human RAD52. Curr. Biol. **18:** 1142–1146.

Ranatunga, W., Jackson, D., Lloyd, J.A., Forget, A.L., Knight, K.L., and Borgstahl, G.E. (2001). Human RAD52 exhibits two modes of self-association. J. Biol. Chem. **276:** 15876–15880.

Rensing, S.A., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319:** 64–69.

Rijkers, T., Van Den Ouweland, J., Morolli, B., Rolink, A.G., Baarends, W.M., Van Sloun, P.P., Lohman, P.H., and Pastink, A. (1998). Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation. Mol. Cell. Biol. **18:** 6423–6429.

Rink, S.M., Lipman, R., Alley, S.C., Hopkins, P.B., and Tomasz, M. (1996). Bending of DNA by the mitomycin C-induced, GpG intrastrand cross-link. Chem. Res. Toxicol. **9:** 382–389.

Rowan, B.A., Oldenburg, D.J., and Bendich, A.J. (2010). RecA main-

tains the integrity of chloroplast DNA molecules in Arabidopsis. J. Exp. Bot. **61:** 2575–2588.

Sadreyev, R., and Grishin, N. (2003). COMPASS: A tool for comparison of multiple protein alignments with assessment of statistical significance. J. Mol. Biol. **326:** 317–336.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. Annu. Rev. Biochem. **77:** 229–257.

San Filippo, J., Chi, P., Sehorn, M.G., Etchin, J., Krejci, L., and Sung, P. (2006). Recombination mediator and Rad51 targeting activities of a human BRCA2 polypeptide. J. Biol. Chem. **281:** 11649–11657.

Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: A new way to display consensus sequences. Nucleic Acids Res. **18:** 6097–6100.

Schuler, G.D., Altschul, S.F., and Lipman, D.J. (1991). A workbench for multiple alignment construction and analysis. Proteins **9:** 180–190.

Shaked, H., Avivi-Ragolsky, N., and Levy, A.A. (2006). Involvement of the Arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. Genetics **173:** 985–994.

Shen, Z., Cloud, K.G., Chen, D.J., and Park, M.S. (1996). Specific interactions between the human RAD51 and RAD52 proteins. J. Biol. Chem. **271:** 148–152.

Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S., and Ogawa, T. (1998). Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. Genes Cells **3:** 145–156.

Shinohara, M., Shita-Yamaguchi, E., Buerstedde, J.M., Shinagawa, H., Ogawa, H., and Shinohara, A. (1997). Characterization of the roles of the Saccharomyces cerevisiae RAD54 gene and a homologue of RAD54, RDH54/TID1, in mitosis and meiosis. Genetics **147:** 1545–1556.

Siaud, N., Dray, E., Gy, I., Gérard, E., Takvorian, N., and Doutriaux, M.P. (2004). Brca2 is involved in meiosis in Arabidopsis thaliana as suggested by its interaction with Dmc1. EMBO J. **23:** 1392–1401.

Singleton, M.R., Wentzell, L.M., Liu, Y., West, S.C., and Wigley, D.B. (2002). Structure of the single-strand annealing domain of human RAD52 protein. Proc. Natl. Acad. Sci. USA **99:** 13492–13497.

Söding, J. (2005). Protein homology detection by HMM-HMM comparison. Bioinformatics **21:** 951–960.

Sonoda, E., Takata, M., Yamashita, Y.M., Morrison, C., and Takeda, S. (2001). Homologous DNA recombination in vertebrate cells. Proc. Natl. Acad. Sci. USA **98:** 8388–8394.

Stasiak, A.Z., Larquet, E., Stasiak, A., Müller, S., Engel, A., Van Dyck, E., West, S.C., and Egelman, E.H. (2000). The human Rad52 protein exists as a heptameric ring. Curr. Biol. **10:** 337–340.

Sun, X., Feng, P., Xu, X., Guo, H., Ma, J., Chi, W., Lin, R., Lu, C., and Zhang, L. (September 20, 2011). A chloroplast envelope-bound PHD transcription factor mediates chloroplast signals to the nucleus. Nat Commun **2** (online), doi/10.1038/ncomms1486.

Swoboda, P., Gal, S., Hohn, B., and Puchta, H. (1994). Intrachromosomal homologous recombination in whole plants. EMBO J. **13:** 484–489.

Symington, L.S. (2002). Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. Microbiol. Mol. Biol. Rev. **66:** 630–670 (table of contents.).

Thorpe, P.H., Marrero, V.A., Savitzky, M.H., Sunjevaric, I., Freeman, T.C., and Rothstein, R. (2006). Cells expressing murine RAD52 splice variants favor sister chromatid repair. Mol. Cell. Biol. **26:** 3752–3763.

Timme, R.E., and Delwiche, C.F. (2010). Uncovering the evolutionary origin of plant molecular processes: Comparison of Coleochaete (Coleochaetales) and Spirogyra (Zygnematales) transcriptomes. BMC Plant Biol. **10:** 96.

Wesley, S.V., et al. (2001). Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J. **27:** 581–590.

Woodson, J.D., and Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. Nat. Rev. Genet. **9:** 383–395.

Wu, S., and Zhang, Y. (2007). LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res. **35:** 3375–3382.

Yoo, S.D., Cho, Y.H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. Nat. Protoc. **2:** 1565–1572.

Zuo, X., Xue, D., Li, N., and Clark-Walker, G.D. (2007). A functional core of the mitochondrial genome maintenance protein Mgm101p in Saccharomyces cerevisiae determined with a temperature-conditional allele. FEM. Yeast Res. **7:** 131–140.